

## Designing an intelligent system for diagnosing type 2 diabetes using the data mining approach: *brief report*

Rohollah Kalhor Ph.D.<sup>1</sup>  
Asghar Mortezaagholi M.Sc.<sup>2</sup>  
Fatemeh Naji M.Sc., M.P.H.<sup>3</sup>  
Saeed Shahsavari M.Sc.<sup>4</sup>  
Mohammad Zakaria Kiaei  
M.Sc.<sup>5\*</sup>

1- Social Determinants of Health  
Research Center, Qazvin University  
of Medical Sciences, Qazvin, Iran.

2- Department of Artificial  
Intelligence, Computer and  
Information Technology (IT)  
Engineering, Faculty of Electrical,  
Qazvin Islamic Azad University,  
Qazvin, Iran.

3- Department of Epidemiology,  
Qazvin University of Medical  
Sciences, Qazvin, Iran.

4- Instructor of Biostatistics, Health  
Products Safety Research Center,  
Qazvin University of Medical  
Sciences, Qazvin, Iran.

5- Instructor in Health Services  
Management, School of Health,  
Qazvin University of Medical  
Sciences, Qazvin, Iran.

\* Corresponding author: Department of  
Health Services Management, School of  
Health, Qazvin University of Medical  
Sciences, Shahid Bahaonar Blvd., Qazvin,  
Iran.  
Post code: 3419759811  
Tel: +98- 28- 33369581  
E-mail: kiaei\_mzsa@yahoo.com

### Abstract

Received: 17 Sep. 2018 Revised: 25 Sep. 2018 Accepted: 22 Feb. 2019 Available online: 01 Mar. 2019

**Background:** Diabetes mellitus has several complications. The Late diagnosis of diabetes in people leads to the spread of complications. Therefore, this study has been done to determine the possibility of predicting diabetes type 2 by using data mining techniques.

**Methods:** This is a descriptive-analytic study that was conducted as a cross-sectional study. The study population included people referring to health centers in Mohammadieh City in Qazvin Province, Iran, from April to June 2015 for screening for diabetes. The 5-step CRISP method was used to implement this study. Data were collected from March 2015 to June 2015. In this study, 1055 persons with complete information were included in the study. Of these, 159 were healthy and 896 were diabetic. A total of 11 characteristics and risk factors were examined, including the age, sex, systolic and diastolic blood pressure, family history of diabetes, BMI, height, weight, waistline, hip circumference and diagnosis. The results obtained by support vector machine (SVM), decision tree (DT) and the k-nearest neighbors algorithm (k-NN) were compared with each other. Data was analyzed using MATLAB® software, version 3.2 (Mathworks Inc., Natick, MA, USA).

**Results:** Data analysis showed that in all criteria, the best results were obtained by decision tree with accuracy (0.96) and precision (0.89). The k-NN methods were followed by accuracy (0.96) and precision (0.83) and support vector machine with accuracy (0.94) and precision (0.85). Also, in this study, decision tree model obtained the highest degree of class accuracy for both diabetes classes and healthy in the analysis of confusion matrix.

**Conclusion:** Based on the results, the decision tree represents the best results in the class of test samples which can be recommended as a model for predicting diabetes type 2 using risk factor data.

**Keywords:** artificial intelligence, data mining, diabetes mellitus type 2, early diagnosis, machine learning, risk factors.